

33 | 服务效率提升：如何降低公司运营成本？

2020-02-10 庄振运

性能工程高手课

[进入课程 >](#)



讲述：庄振运

时长 15:36 大小 12.51M



你好，我是庄振运。

我们都知道，支持大量用户的互联网公司，通常会部署相当规模的系统容量来运行各种服务。

如果你想要有效地运行业务，就应使业务的**容量需求**和**容量供应**尽可能地相等。为什么这么说呢？如果容量供应不能满足需求，那么部分业务将因容量不足，不能部署或扩展。如果容量供应过多，那么公司基础设施的效率就降低了。



服务效率与业务的容量和公司预算直接相关。因此，**提高服务效率**和**适当地预测容量需求**，对于公司的持续成功至关重要。

那么如何才能降低公司的服务容量需求呢？这就需要**运行的服务尽量高效**，提高服务效率将会帮助降低容量的需求。

今天我就为你介绍八种提升服务效率的途径和相关的生产经验，以及提升执行服务效率的原则，希望你有所启发。

服务效率提升的途径和类别

当各个服务团队的领导和技术骨干下定决心，要花时间在服务效率的时候，经常提出的问题是：**怎么样做才能提高服务效率呢？**

要回答这个问题，你就必须先搞清楚一个问题，就是——什么类型的工作才是提升服务效率的工作？

我多年来帮助很多不同的服务提升了服务效率；在这个过程中我逐渐意识到，一个公司的服务千差万别，可以提升效率的方式也是多种多样。但核心的一点是：任何能够让这个服务降低容量需求的工作，都是服务效率提升的工作。

我把各种各样的服务效率提升工作分为八类：软件效率（Software Efficiency）、服务架构效率（Service Architecture Efficiency）、服务部署效率（Service Deployment Efficiency）、跨平台效率（Cross-platform Efficiency）、硬件效率（Hardware Efficiency）、容量组织效率（Capacity Organization Efficiency）、容量资源回收（Capacity Resource Reclamation）、用户效率（User Efficiency）等。

这八种效率大体上分为两组：**软件服务**和**非软件容量**。

软件服务

软件效率

服务架构效率

服务部署效率

跨平台效率

非软件容量

硬件效率

容量组织效率

容量资源回收

用户效率



对每一种服务效率，我先讲基本概念，然后再用具体例子来诠释，希望可以给你些启发。

1. 软件效率

服务的软件程序实现千差万别，但不管哪种实现，它总是可以优化的，比如通过采用更好的数据结构和语言库等等。这方面的效率提升，都可以归类为软件效率。

软件方面的效率提升比较直白，就是把程序的性能提高。比如减少 CPU 和内存的使用等等；这方面我们这个专栏前面讲了很多，多数内容都可以归到这方面。

2. 服务架构效率

对一种服务而言，如果能够改进服务的实现架构，比如重新整合了内部的微服务，从而变得更高效，那么就是服务架构的效率提升。

服务架构的效率又可以分为两种：一种是单独的一个服务，通过进行服务设计的优化来提高效率；另外一种就是通过合理借助其他服务来优化。

比如一个数据库服务，如果有很多的查询请求，那么采用另外一个缓存服务，就或许可以大幅度地降低数据库服务的资源消耗。如果两个服务（数据库和缓存）的资源使用总和，还是小于仅仅使用数据库（而没有缓存服务）的资源使用，那么这两种服务的结合就是一种更加高效的服务架构。

3. 服务部署效率

一个服务的软件程序总是要部署到硬件容量上。生产实践中，如何部署软件也很重要，也会对服务效率产生影响。比如采用什么操作系统，进行什么样的系统和软件配置，要不要采用 NUMA 绑定等。这些方面的优化都可以算是服务部署方面的效率提升。

还记得我们在 [第 22 讲](#) 讲过的“使用内存大页面，来装载程序的热文本区域”吗？这其实就是一项服务部署的优化，就是通过提高系统和程序 iTLB 命中率，从而达到更高的服务效率。

4. 跨平台效率

如果一个公司（尤其是后端）同时提供几种服务平台，提供的功能有重合，而且服务效率不同，那么用户可以迁移到效率更高的服务上去。这种迁移就是跨平台效率提升。

一个互联网公司，随着时间的推移和规模的增长，内部往往有很多种服务提供重叠的功能。一个明显的例子是数据存储，经常有很多服务都可以提供数据存储的功能。虽然每种存储服务提供的功能并不完全一样，但对于一个要使用存储服务的用户来说，经常同时会有好几个选择。

这种情况下，这个需要存储的用户，就可以考虑每种存储服务的服务效率；在满足所需功能的前提下，选择一种高效的服务，会帮助公司降低成本。

对于一个已经在使用某种服务的用户而言，可以重新考虑各种可用服务的效率；如果另外一种服务更加高效，就可以考虑迁移过去。这种操作就是**跨平台的迁移**，其结果也是能够提升公司的服务效率的。

我最近几年做过很多这方面的优化。比如曾经把一个用户的数据从一个低效的存储服务，整体转移到另外一个高效的存储服务，从而大幅度地降低了公司的容量和运营成本，节省了大约几千万美元。重要的是，用户的端到端性能并没有收到影响。

5. 硬件效率

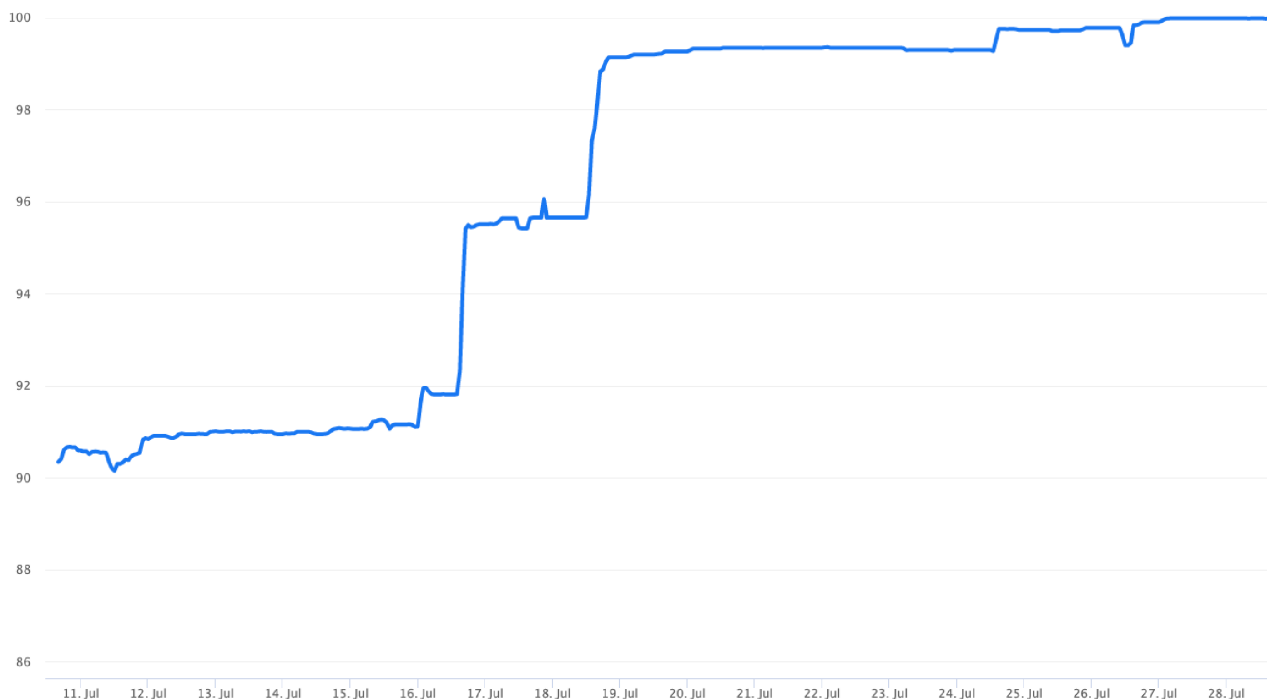
一个服务会使用各种容量资源（比如 CPU）。所以，根据资源使用最大瓶颈的不同，采用不同种的服务器硬件，也会导致不同的服务效率。比如，假设一个服务的存储是最大瓶颈时，那么采用相对较大存储的服务器，结果就会比较高效。

我就提升过一个存储服务的硬件效率。该服务受到存储空间的限制，并且近年来增长迅速。我评估后得出的结论是，如果这个服务继续快速地增长，公司很快就没有足够的容量来提供给它。

所以我们决定，通过提升服务效率来减少服务的容量需求。在检查了所有效率提升的方式后，我决定首先提升硬件效率。具体来说，我们采用了一种新型服务器，这种服务器相对旧的服务器而言，有较大存储容量，并逐步部署。

下图分别显示了服务的容量规模（也就是服务器的数目），和已部署的总存储容量。





在 20 天的时间内，我们将服务迁移到新的硬件类型，服务器的数目减少了 5%，而存储容量增加了 11%。值得一提的是，新旧服务器的单位服务器成本差别不大，所以整个服务的效率得到了很好的提升。

6. 容量组织效率

一个互联网服务在使用数据中心提供的容量时候，一般都是用某种方式把容量（比如服务器）组织在一起，形成一个容量单位。比如把服务器分组、组成集群等。这种容量组织的方式，也可以根据服务的特点，进行调整优化，从而变得更高效，这就是容量组织效率。

一个规模很大的服务可能会需要很多容量。我曾经合作的一个后台服务，就使用了几十万台服务器。这么多服务器经常会分成几个容量单元，这样做当然有利有弊，但都是经过各种考虑的决定。有时候服务容量的分割也有历史的原因，比如一开始使用一个容量单元，但是随着服务的规模不断扩大，就建立了一个又一个的容量单元。

分成几个容量单元的一个好处，是降低了容量操作的复杂度。不过也有坏处，其中之一就是资源效率使用率不会特别高，因为不同容量之间的资源不能互补。比如两个容量单元，每个有一千台服务器。一个容量单元是 CPU 吃紧，另外一个容量单元是内存吃紧。如果二者合并成一个容量单元，或许只需要一千五百台服务器就够了。

所以适度地优化容量组织，合并容量单元，就可以降低容量的需求，提高容量的效率。

7. 容量资源回收

一个服务所使用的容量，随着时间的变化，有些容量（比如一些服务器）就会处于空闲状态。如果及时回收，从而进行容量的再利用，也可以提升服务效率。

一个服务从诞生到成熟，会不断地演化；表现在对容量的需求方面，也是不断变化的。可能有一段时期，部署的容量恰好满足服务的需求，容量的效率很高。但是过一段时间，服务的需求和资源使用特征发送了变化，有些容量和相对应的资源使用不再使用。这种情况下就需要进行容量和资源回收。

这方面一个最直白的表现，就是可能有些容量处在空闲状态。举几个例子，假如有些服务器已经不能运行此种服务了，那么就把这些服务器回收，给其他合适的服务用。又假如一个服务对存储资源的使用下降了，那么对应的存储资源就可以减少，分给别的服务用。

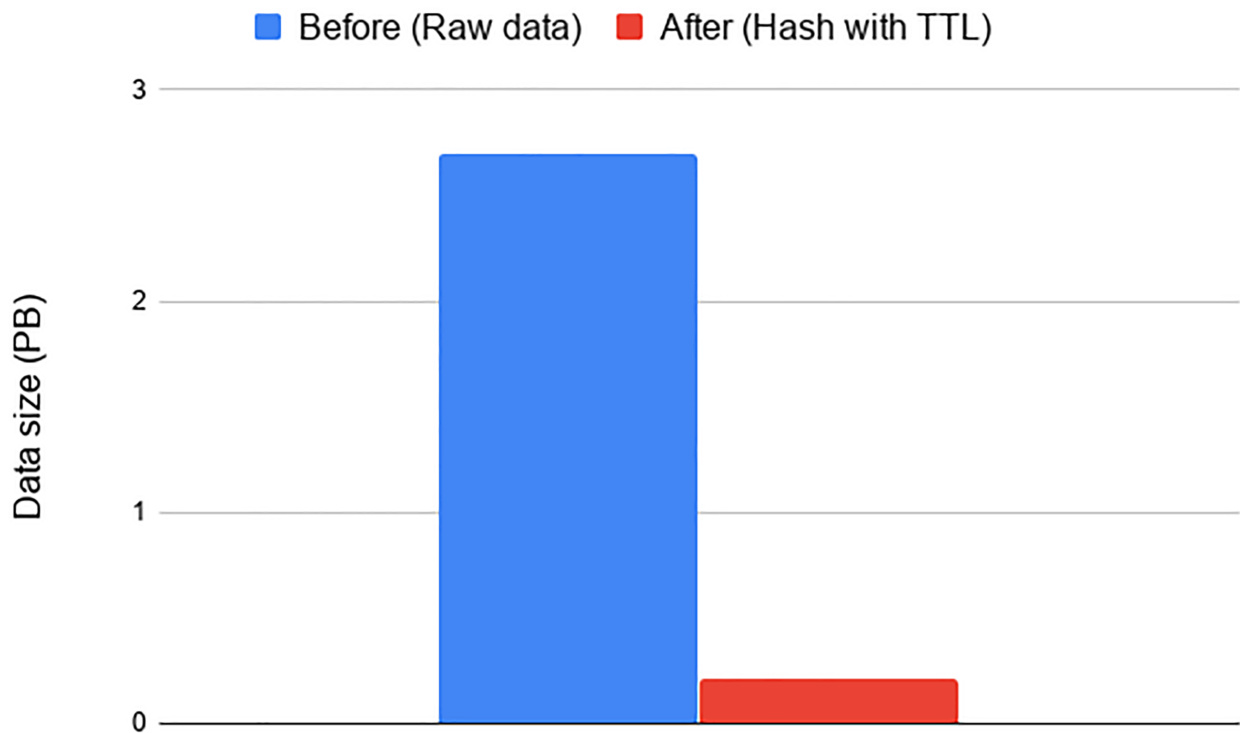
8. 用户效率

使用服务的用户，如果采用合理的方式，也可以帮助提升所使用服务的效率。这方面的效率提升就是用户效率。

用户在使用服务时，可以尽量做到高效使用。我们曾经帮助一个存储服务提升效率，通过和客户合作，大幅度降低了客户的存储数据量。具体来说，这个客户以前是将数据对象的本来数据和一大堆元数据，都放到了存储服务上，这导致了相当大的物理数据存储。

我们是怎么提高用户效率的呢？我们首先分析了客户的业务逻辑和数据使用场景，意识到客户存储的目的，是判断数据对象是否已经更改。我们先将存储的数据分为两部分，然后分别存储对象的哈希（例如 MD5），而不是对象本身。另外，我们把数据的 TTL（生存时间）缩短，并在对象哈希中定期强制清除陈旧数据。

这样的优化工作取得了很好的效果，节省了 93% 的存储空间。具体来说，把客户数据大小从 2.7PB，降到了 0.2PB，如下图所示。



执行服务效率提升的原则

除了前面说过的八类服务效率提升方法，在提高服务效率的生产实践中，我也获得了许多执行方面的经验。这些经验对于如何在部门之间协调，有效地执行效率提升，完成预定的效率提升目标等方面，有比较好的指导作用，希望能对你有所帮助。

1. 明确的所有权和承诺。

根据公司的规模，公司可能会有许多需要提高效率的服务。对于每个这样的服务，每个效率提升工作都需要有明确的拥有者团队，并且需要得到团队的工作承诺。如果不能明确所有者，或者所有者不能承诺完成的时间，那么最后经常会互相扯皮，互相指责，工作失败。

2. 以效率目标为导向的计划和执行。

对于每个服务及其效率组件，要定期设定要实现的效率目标。例如，针对存储服务的效率目标可以是：“将明年的存储空间利用率从 30% 提高到 40%”。但是注意，目标设定的频率和大小，必须适应服务的特征和团队的工作特点。

3. 为每个效率组件部署检测工具。

你需要开发一定的工具，来监视整个服务和每个组件的效率水平，以便快速检测到异常。对于异常的效率降低，这些工具必须有自动触发或警报功能。有效的工具，能帮助你及时发现问题，并立即采取行动。

4. 紧密的跨团队协作。

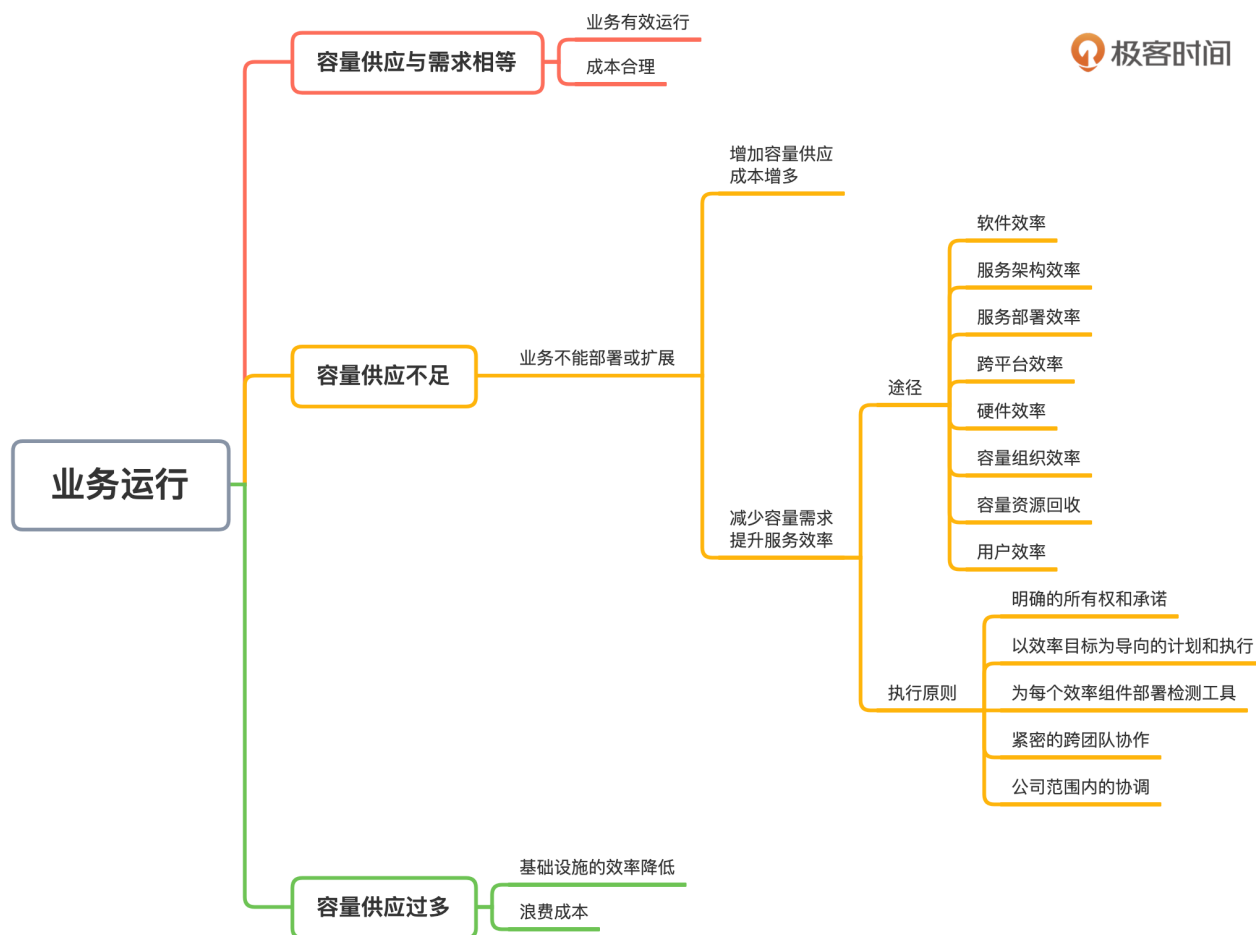
鉴于当今业务和服务的复杂性，许多服务效率的提高，都需要多个团队共同合作。例如，某个服务可能依赖其他服务，并且该服务的效率提高，也可能需要所依赖服务团队的协作。

5. 公司范围内的协调。

尽管每个主要的独立服务，都可以通过采用我们提出的框架来独立工作，但也需要在公司范围内进行协调，因为公司需要保证总体业务的需求。

总结

互联网公司业务复杂，规模庞大，需要有大量的基础设施和容量去支撑。我们这一讲讨论了如何提升服务效率，降低容量需求，从而节省公司成本。



通常一个互联网服务的规模是不断增长的。开始的时候，或许你不在乎它的服务效率和系统容量，但是很快你就需要重视了。所以，你最好未雨绸缪，把相关工作定义好。正如唐朝诗人杜荀鹤的一首《小松》，有两句说：“时人不识凌云木，直待凌云始道高。”如果公司业务发展快，它的服务规模增长也很快，那么服务效率提升的工作就会越发重要。

我通过多年的容量规划和服务效率提升工作，积累了大量的服务效率优化的知识，今天的分享，希望能帮你开拓思路，提供经验。

思考题

提升一个互联网服务的效率有很多种方式，其中一个是服务软件的效率。我们前面几讲提到了很多性能优化的工作。你能举出几个可以归类于软件效率提升的例子吗？提示一下，优化代码算不算是提升服务软件的效率？

欢迎你在留言区分享自己的思考，与我和其他同学一起讨论，也欢迎你把文章分享给自己的朋友。

点击查看 

21天打卡, Facebook 工程师 带你攻克性能工程



PC 端用户扫码参与



新版升级：点击「 请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 32 | 服务的容量规划：怎样才能做到有备无患？

精选留言

 写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。